

# Longitudinal Study of EITC Claimants

*Karen Masken, Internal Revenue Service*

---

**T**he Earned Income Tax Credit (EITC), enacted in 1975, provides a refundable tax credit for low-income working families. Originally intended to ease the burden of Social Security taxes and provide an incentive to work, the credit has been modified several times during the years since its introduction. The credit now provides a substantial benefit to millions of American taxpayers.

While it is known that there is significant turnover in EITC claimants from one year to the next, the reasons for this are not well understood. In order to better understand why taxpayers move in and out of the EITC population, the Office of Research is conducting a longitudinal study of tax returns filed for Tax Years 1996 through 2004. In addition to tracking taxpayers who claimed EITC in at least 1 of the last 9 years, the study will also track the children claimed in the last 4 years (due to data problems, it is not possible at this time to track the children for all 9 years). This paper presents some of the data issues encountered and a preliminary analysis of taxpayer patterns during the study period. It also looks at the pattern of children claimed as qualifying children for the shorter time period.

## **Methodology**

The study is based on administrative data stored in the Compliance Data Warehouse (CDW) and includes the entire population of EITC claimants for Tax Years 1996-2004 that were processed through 2005. Typically, when the IRS refers to an individual taxpayer, the reference is to one Form 1040 return. In more general terms, the Form 1040 return can be thought of as a household comprised of the primary and secondary taxpayers along with their dependents. It is generally accepted that trying to follow a household over time becomes virtually impossible due to constant changes in household composition. Therefore, this study follows individual persons (about 70 million taxpayers and 28 million children), not returns. For example, if a married couple files a joint tax return and claims the EITC with two qualifying children, then both the primary and secondary taxpayers are followed as well as both of the children.

## **Data Source**

As mentioned above, the file is based on population data stored in CDW. There are several advantages to using this administrative population data. First, it allows for a longitudinal file to be built retrospectively. Also, since it is not based on a sample, it is not dependent on any underlying sample design. This is particularly important when there are changes in tax law since a sample may not adequately capture or reflect responses to tax law changes. Finally, it allows for individuals to be followed. The ability to follow both the primary and secondary taxpayers alleviates several issues encountered with sample panel data in which only the primary taxpayer is followed. Following only the primary taxpayer can lead to false attrition rates when the couple stops filing a joint return and the secondary taxpayer continues to claim the EITC while the primary taxpayer does not. In this instance, sample data would not capture the behavior of the secondary taxpayer. This also leads to gender bias over time since the secondary taxpayer is typically female. Using this population data makes it possible to capture changes in the composition of the household and follow all members of the household.

## **Data Issues**

### **Multiple Returns for 1 Tax Year**

On average, there were approximately 1.2 million duplicate or multiple returns filed each year. In cases where a taxpayer filed multiple, different returns for the same tax year, the return with the latest tax period and highest EITC claim was selected. (The tax period refers to both the tax year and the last month in the accounting year. While most taxpayers file on a calendar-year basis, there are some who file on another basis, such as fiscal year.) Duplicates returns were simply removed. There were also about 220,000 returns each year where the person being followed was a secondary taxpayer on more than one return. Again, the return with the latest tax period and highest EITC claim was selected. In cases where the person was listed as a primary on one return and a secondary on another return for the same Tax Year, the return where they were listed as a primary taxpayer was selected.

### **Missing and Incomplete Data**

It appears that Tax Year 1999 is missing about five million returns and, as a consequence, return information for approximately 1.7 million people in the study is missing. Also, about three-quarters of one million EITC claims are made in later years, so that the Tax Year 2004 information is incomplete. While

this introduces some noise into the data, it is still valuable to look across all 9 years.

Data for the children are incomplete for tax years prior to 2001, and, therefore, the analysis for the children can only be conducted for Tax Years 2001-2004. Again, 2004 is incomplete due to late filers. There are also several suspect child Taxpayer Identification Numbers (TINs) used by a large number of children (for example, children with the TIN 123-45-6789 appears more than 10,000 times on the files). The reasons for this are not well understood, and they have been excluded from this analysis.

### **Unedited Data Fields**

The administrative data have two fields for the amount of EITC claimed. One field is “per taxpayer” which is ostensibly what the taxpayer reported on his or her return. The other is “per computer” which is the IRS computed amount. In theory, these two fields should differ only if there is an EITC-related math error. However, the “per taxpayer” field also contains transcription errors--some of which are quite large (\$97 million was the largest, the actual maximum is about \$4 thousand).

Because the number of math errors has declined over time, it is not appropriate to compare the “per computer” amounts across time when attempting to understand taxpayer behavior. The “per taxpayer” is the appropriate field and an attempt was made to clean up the transcription errors systematically. All claims were capped at the maximum EITC allowed for the given tax year. Also, lagging zeroes were checked for, and, finally, if there did not appear to be a math error, the “per taxpayer” was set to “per computer.”

## **Analysis**

### **General Trends**

Figure 1 presents the amount of EITC claimed over time in real 2004 dollars (the CPI was used as the inflator). Due to noise in the data discussed previously, the drop in Tax Year 1999 is probably overstated; however, the downward trend at a time the economy was strong is likely accurate. The jump in 2002 is due to several tax law changes. Since Tax Year 2004 is incomplete, it is not included in this graph.

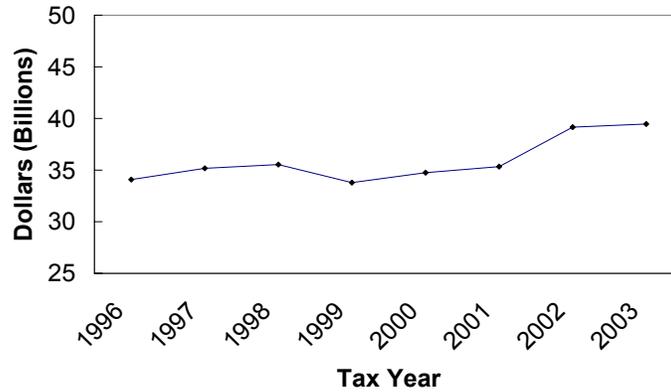
**Figure 1. EITC Claims in 2004 Dollars**

Figure 2 presents the percentage of all individual taxpayers claiming EITC in each tax year. As would be expected, the percentage dropped when the economy was strong, and then started climbing as the economy weakened. Also, the tax law change in 2002 increased the percentage of taxpayers claiming EITC.

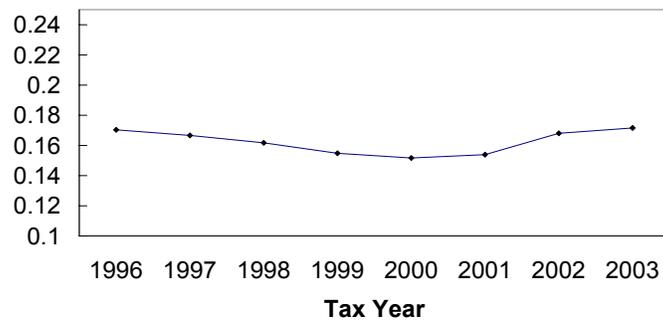
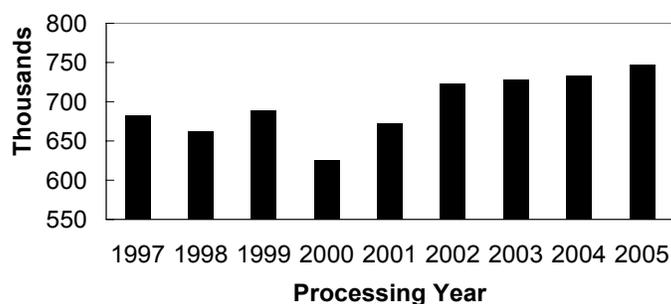
**Figure 2. Percent of Individual Taxpayers Who Claim EITC**

Figure 3 shows the number of returns each processing year with EITC claims for prior tax years. For example, in Processing Year 2005, there were approximately 750,000 returns with claims for Tax Year 2003 or before. (The drop in 2000 is likely overstated due to the Tax Year 1999 data issue already discussed.)

**Figure 3. Number of Taxpayers Each Processing Year Who File Claims for Prior Tax Years**



**Taxpayer Patterns**

Table 1 portrays the most frequent filing patterns for individuals in the study. Each column represents a tax year (beginning with Tax Year 1996). Thus, an ‘X’ in the first column indicates that a return was filed for Tax Year 1996, while a dash indicates one was not. As shown, the plurality (47 percent) of people in the study filed a return in each of the 9 years studied. These 17 patterns displayed in the table (of a possible 511 patterns) account for 75 percent of the study population. The fourth-row pattern is due to data problems with Tax Year 1999. It is likely that the majority of people in this category actually belong in the first-row category. Aside from this issue, it is interesting to note that the majority of people in the study do not file sporadically. Once they file, they continue to file, and, once they stop filing, they do not re-enter the filing population.

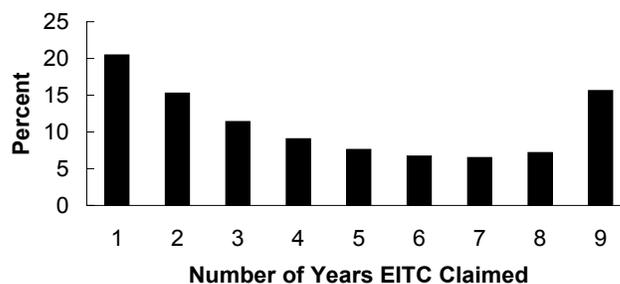
Table 2 shows the most frequent patterns for claiming the EITC. These 18 patterns (again, there are 511 possible patterns) account for about 50 percent of the population. Approximately 7 percent of individuals in the study claim the

EITC persistently. It is interesting to note that, much like the filing patterns, the most frequent patterns of claims are not sporadic.

While the above patterns are interesting, they are confounded by nonfilers since claiming the EITC is dependent on filing a return. Figure 4 shows the number of years EITC was claimed by individuals who filed returns in each of the 9 study years. A little over 20 percent claimed EITC in only 1 year, while slightly over 15 percent claimed it in all years.

Table 3 presents the most frequent pattern of claims for study members who filed returns in each of the 9 years. Like the overall patterns, individuals do not appear to move in and out of the claimant population sporadically.

**Figure 4. Number of Years EITC Claimed by Individuals Who Filed All Nine Years**



### Qualifying Child Patterns

As mentioned earlier, only Tax Years 2001-2004 can be analyzed for the qualifying children due to data constraints. The children included in the study are children who were claimed at least once as a qualifying child in this time frame. In order to be claimed as a qualifying child for EITC, the child must meet certain age, relationship, and residency tests. A child who meets these qualifying child requirements could also meet the requirements to be claimed as a dependent, but this is not necessarily so. It is possible for a child to be claimed correctly by one taxpayer as a dependent and by another as a qualifying child. The first column in Table 4 displays all possible patterns of children being claimed either as a dependent (second column) or as a qualifying child (third column) during the 4-year study period. For children who were claimed as dependents on only one return in any given year (95 percent of the children in the study), 60 percent were claimed every year as dependents. In comparison, for children claimed on only one return in any given year as a qualifying child

(98 percent of children in the study), 31 percent were claimed every year as a qualifying child. Interestingly, about one-half of 1 percent were never claimed as dependents but were claimed as qualifying children for EITC.

Of those being claimed as qualifying children in each of the 4 years, 75 percent were consistently claimed as both a dependent and as a qualifying child by the same primary taxpayer in each year. However, a large number (21 percent), were claimed as both a dependent and qualifying child in each year, but not by the same taxpayer across years. Table 5 illustrates the number and pattern of taxpayers claiming the child as a qualifying child across the years. Each number in the pattern column represents a different taxpayer. For example, the pattern '1 2 1 2' indicates two different taxpayers claiming the child in alternating years, whereas the pattern '1 2 3 4' indicates the child was claimed by a different taxpayer every year.

### Next Steps

While this analysis provides valuable insight into *what* taxpayers do, the primary goal of conducting a longitudinal study is to try and understand *why* taxpayers move in and out of the EITC claimant population. Future research will try to understand from the administrative data why taxpayers enter and why they leave the claimant population. It is also hoped that more retrospective years can be obtained for the children in order to better understand the patterns that exist. It is also of interest to try to understand why some children are claimed by more than one taxpayer, particularly in one given year.

Filing Pattern	Percent	Cumulative Percent
X X X X X X X X X	47%	47%
- X X X X X X X X	4%	51%
X X X X X X X X -	3%	54%
X X X - X X X X X	3%	56%
- - X X X X X X X	2%	59%
- - - X X X X X X	2%	61%
- - - - X X X X X	2%	63%
X X X X X X X - -	2%	65%
- - - - - X X X X	2%	66%
X X X X X X - - -	1%	67%
- - - - - X X X	1%	69%
X X X X X - - - -	1%	70%
X X X - - - - - -	1%	71%
X - - - - - - - -	1%	72%
- - - - - - X X	1%	73%
X X - - - - - - -	1%	74%
X X X X - - - - -	1%	75%

Table 2. Most Frequent Patterns of Claiming EITC										
Claims Pattern									Percent	Cumulative Percent
<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	7%	7%
<b>X</b>	-	-	-	-	-	-	-	-	6%	13%
-	-	-	-	-	-	-	<b>X</b>	<b>X</b>	3%	17%
-	-	-	-	-	-	-	<b>X</b>	-	3%	20%
<b>X</b>	<b>X</b>	-	-	-	-	-	-	-	3%	24%
-	-	-	-	-	-	<b>X</b>	-	-	3%	27%
-	<b>X</b>	-	-	-	-	-	-	-	3%	29%
-	-	-	-	-	-	<b>X</b>	<b>X</b>	<b>X</b>	3%	32%
-	-	<b>X</b>	-	-	-	-	-	-	2%	35%
<b>X</b>	<b>X</b>	<b>X</b>	-	-	-	-	-	-	2%	37%
-	-	-	-	-	<b>X</b>	-	-	-	2%	39%
-	-	-	-	-	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	2%	41%
-	-	-	-	<b>X</b>	-	-	-	-	2%	43%
-	-	-	<b>X</b>	-	-	-	-	-	2%	45%
-	-	-	-	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	2%	46%
<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	-	-	-	-	-	2%	48%
-	<b>X</b>	2%	49%							
-	-	-	-	-	-	<b>X</b>	<b>X</b>	-	1%	51%

Table 3. Most Frequent Patterns of Claiming EITC by Individuals Who Filed all Nine Years								
Claims Pattern	Percent	Cumulative Percent						
X X X X X X X X X	16%	16%						
X - - - - - - - -	7%	22%						
X X - - - - - - -	4%	26%						
- - - - - - - X -	3%	29%						
- X - - - - - - -	3%	31%						
X X X - - - - - -	3%	34%						
- - - - - - - X X	2%	37%						
- - - - - - - X - -	2%	39%						
- - - - - - - X X X	2%	41%						
X X X X - - - - -	2%	43%						
- - X - - - - - -	2%	45%						
X X X X X X X X -	2%	47%						
- X X X X X X X X	2%	48%						
- - - X - - - - -	1%	50%						
X X X X X - - - -	1%	51%						

Table 4. Pattern of Children Claimed as:				
Claim				
Pattern	Dependent	Qualifying Child		
X X X X	60%	31%		
- X X X	6%	8%		
- - X X	5%	8%		
X X X -	5%	7%		
- - - X	5%	10%		
X X - -	3%	6%		
X - - -	3%	8%		
X X - X	2%	2%		
X - X X	1%	2%		
- - X -	1%	4%		
- X X -	1%	3%		
- X - -	1%	5%		
X - - X	1%	1%		
- - - -	1%	n.a.		
- X - X	*	1%		
X - X -	*	1%		

\* Less than 0.5%

Pattern	Percent	Cumulative Percent
1 1 1 1	77%	77%
1 2 2 2	6%	83%
1 1 1 2	5%	88%
1 1 2 2	4%	92%
1 1 2 1	1%	93%
1 2 1 1	1%	95%
1 1 2 3	1%	96%
1 2 3 3	1%	97%
1 2 2 3	1%	98%
1 2 3 4	1%	99%
1 2 2 1	1%	99%
1 2 1 2	*	99%
1 2 1 3	*	100%
1 2 3 1	*	100%
1 2 3 2	*	100%

\* Less than 0.5%